Masked Sentence Model based on BERT for Move Recognition in Medical Scientific Abstracts

Gaihong Yu [1,2], Zhixiong Zhang*[1,2,3], Huan Liu [1,2], Liangping Ding[1,2]

(1. National Science Library, Chinese Academy of Sciences, Beijing 100190
2.University of Chinese academy of sciences, Beijing 100049, China
3.Wuhan Library, Chinese Academy of Sciences, Wuhan 430071,China *)*

Abstract:

Purpose: Move recognition in scientific abstracts is an NLP task of classifying sentences of the abstracts into different types of language unit. To improve the performance of move recognition in scientific abstracts, a novel model of move recognition is proposed that outperforms BERT-Base method.

Design: Prevalent models based on BERT for sentence classification often classify sentences without considering the context of the sentences. In this paper, inspired by the BERT's Masked Language Model (MLM), we propose a novel model called Masked Sentence Model that integrates the content and contextual information of the sentences in move recognition. Experiments are conducted on the benchmark dataset PubMed 20K RCT in three steps. And then compare our model with HSLN-RNN, BERT-Base and SciBERT using the same dataset.

Findings: Compared with BERT-Base and SciBERT model, the F1 score of our model outperforms them by 4.96% and 4.34% respectively, which shows the feasibility and effectiveness of the novel model and the result of our model comes closest to the state-of-the-art results of HSLN-RNN at present.

Research Limitations: The sequential features of move labels are not considered, which might be one of the reasons why HSLN-RNN has better performance. And our model is restricted to dealing with bio-medical English literature because we use dataset from PubMed which is a typical bio-medical database to fine-tune our model.

Practical implications: The proposed model is better and simpler in identifying move structure in scientific abstracts, and is worthy for text classification experiments to capture contextual features of sentences.

Originality: The study proposes a Masked Sentence Model based on BERT which takes account of the contextual features of the sentences in abstracts in a new way. And the performance of this classification model is significantly improved by rebuilding the input layer without changing the structure of neural networks.

Keywords: Move Recognition, BERT, Masked Sentence Model, Scientific Abstracts

1. Introduction

* Corresponding author: Zhixiong Zhang (OCRID: 0000-0003-1596-7487, E-mail: zhangzhx@mail.las.ac.cn)

The concept of move, or rhetorical move, was originally developed by Swales to functionally describe a part or section in research articles for communicative purpose (Swales et al., 2004). Authors of research papers generally need to explain the purpose, methods, results, and conclusions of their researches in abstracts. Those language units are called as the moves of the abstracts.

Many journals currently require authors to provide structured abstracts with explicitly annotated move labels. For example, authors usually use “Purpose” to indicate the move label and the sentences following the label to represent the aim of the study. But currently, many important journals such as Nature and Science still use unstructured abstracts when they publish the research articles.

Automatically recognizing moves of unstructured abstracts in research papers(move recognition in brief), which is typically a classification task, enables readers to quickly grasp the main points of research papers, and it is useful for various text-mining tasks such as information extraction, information retrieval and automatic summarization.

Many researchers have done lots of work on it. Early researches on move recognition adopted traditional machine learning methods such as Native Bayes (NB) (Teufel,1999), Conditional Random Fields (CRF) (Hirohata et al.,2008), Support Vector Machine (SVM) (Yamamoto and Takagi, 2005; Ding et al.,2019), Logic Regression (LR) (Fisas et al.,2016) etc. These methods achieve good recognition performance, but they are very complicated to apply because they rely heavily on numerous carefully hand-engineered features such as lexical, semantic, structural, statistical and sequential features.

In recent years, neural networks have been widely used in NLP research including move recognition tasks (Dasigi et al., 2017; Kim, 2014; Lai et al., 2015; Ma et al., 2015; Zhang et al., 2019). Neural networks have strong nonlinear fitting ability and can automatically learn a better and deeper presentation for the input without complicated feature engineering. Methods using neural networks usually get better performance than traditional machine learning methods, which is one of the reasons why deep learning methods are widely used in many NLP researches.

Especially, Di Jin et al. (Jin and Szolovits, 2018) from MIT proposed a hierarchical sequential labeling network named HSLN-RNN which used the contextual information within surrounding sentences to help classify the current sentence. Concretely, HSLN-RNN used a Bi-LSTM layer after encoding sentence-level features to capture contextual features within sentences and a CRF layer to capture sequential features within surrounding move labels. And HSLN-RNN achieved the state-of-the-art results with a F1 score 92.6% on the dataset of PubMed 20K RCT.

BERT (Bidirectional Encoder Representations from Transformers) (Devlin et al., 2018), released by Google in Oct.2018, received widespread attention because it broke the records of 11 NLP tasks when released. After the release of BERT, some researchers have done move recognition researches based on BERT and try to get better performance. Iz Beltagy et al. (Beltagy et al.,2018) fine-tuned BERT-Base model on PubMed 20K RCT and got an average F1 score of 86.19%. They also released the SciBERT model which re-pre-trained the original BERT model with corpus from biomedical domain. Based on the SciBERT, the F1 score on PubMed 20K RCT reached to 86.81%

which was better than BERT-base model but still didn't reach to the highest F1 score of 92.6% based on HSLN-RNN model (Jin and Szolovits, 2018).

By comparing the models based on BERT and HSLN-RNN, we find that the main problem of the current BERT-based models is that they only use the content information of the sentence without considering the context information of the sentence. Here, "content information" means the sentence itself and the "context information" means the surrounding information (or the surrounding sentences) of the sentence in an abstract.

We assume that the move type of a sentence depends on not only the sentence itself but also its surrounding sentences and the context information of the sentences can help to improve the performance of move recognition. For instance, when we draft an abstract, the sentence of "Results" is more likely to be followed by a sentence of "Conclusions" than a sentence of "Purpose".

In our study, we intend to integrate the content and context information of sentences in move recognition based on BERT. Inspired by BERT's "masked language model" (MLM), we propose a "masked sentence model" (MSM) based on BERT to solve this problem. We mainly improve the move recognition task during the BERT fine-tuning procedure without changing its neutral networks. The model makes full use of the content and context of the sentences.

Our key contributions are summarized as follows:

(1) We propose a Masked Sentence Model based on BERT that can capture not only the content features but also contextual features of the sentences. And our model is easy to apply because it just rebuilds the input layer without any change on the structure of neural networks.

(2) We evaluate on the public dataset for move recognition (PubMed 20K RCT) and see an improvement of approximately +4.34% F1 score compared to SciBERT, +4.96% F1 score compared to BERT-Base model, which shows the effectiveness of our masked model.

2. Methodology

2.1 Main Idea

Firth (Firth,1957) proposed a distributional hypothesis in natural language processing research that words can be identified by their context. This hypothesis has been widely used in information retrieval, topic recognition (Basili and Pennacchiotti,2016) and other NLP researches. BERT's Masked Language Model (MLM) is also based on this distributional hypothesis. MLM simply masks some percentage of the input tokens at random, and then predicts those masked tokens based on their context.

Similarly, we propose that sentences in an abstract also follow the distributional hypothesis and we believe that a sentence in an abstract can be identified by the contextual sentences surrounding them.

Based on this hypothesis, a novel model called Masked Sentence Model (MSM) is proposed. In this model, it integrates two ideas of sentence representations for the move recognition task. Same as traditional deep learning classifiers, it preserves using the content of the target sentence as the input of the classifier to learn the internal features of this sentence. Moreover, for the purpose of capturing

the contextual information, it innovatively uses the whole abstract but with the target sentence masked as the input of the classifier to learn the contextual features of the target sentence.

For example, there is an abstract document from PubMed[1] shown in figure 1 which contains seven sentences (from s_1 to s_7). For the second sentence(s_2), we have two representations for the input of a deep learning classifier: 2-a (representation based on the content of the sentence), 2-b (representation based on the context of the sentence by using the whole abstract masking the target sentence with a fixed meaningless string denoted as '[MASK]'). And in the Masked Sentence Model, we combine the two representations above to learn both the content features and contextual features of the sentence(2-c).

[s_1]This survey aims at reviewing the literature related to Clinical Information Systems (CIS), Hospital Information Systems (HIS), Electronic Health Record (EHR) systems, and how collected data can be analyzed by Artificial Intelligence (AI) techniques. **[s_2]*We selected the major journals (11 journals) collecting papers (more than 7,000) over the last five years from the top members of the research community, and read and analyzed the papers (more than 200) covering the topics.*** [s_3]Then, we completed the analysis using search engines to also include papers from major conferences over the same five years. [s_4]We defined a taxonomy of major features and research areas of CIS, HIS, EHR systems. [s_5]We also defined a taxonomy for the use of Artificial Intelligence (AI) techniques on healthcare data. [s_6]In the light of these taxonomies, we report on the most relevant papers from the literature. [s_7]We highlighted some major research directions and issues which seem to be promising and to need further investigations over a medium- or long-term period.

Figure 1 An example of an abstract

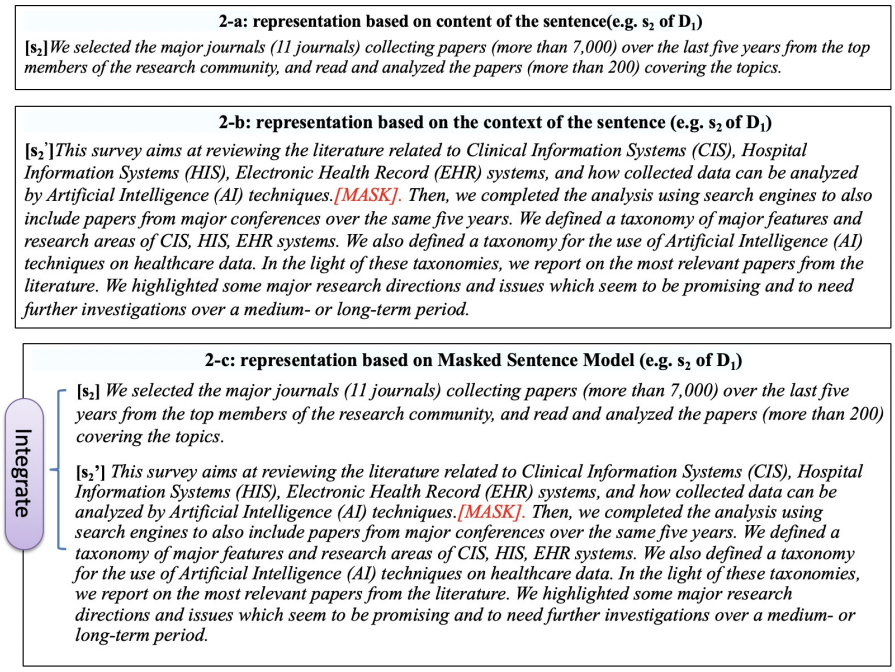

Figure 2 Sentence representations

We use this integrated MSM representation of the sentence as input in the BERT fine-tuning procedure and conduct several experiments to verify its effectiveness.

2.2 MSM Construction

Based on the main idea mentioned above, we construct the Masked Sentence Model (as shown in figure3) in three processing steps before BERT fine-tuning.

1 https://www.ncbi.nlm.nih.gov/pubmed/31419820

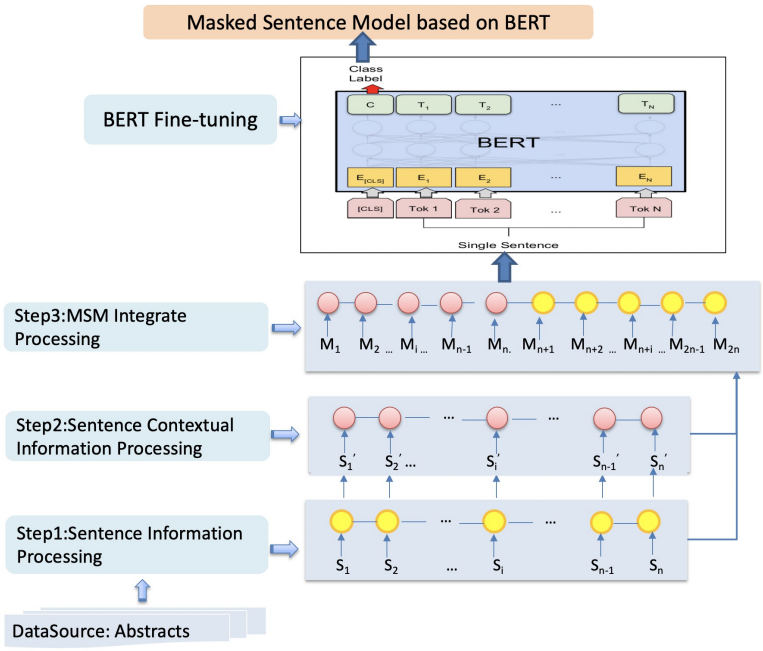

Figure 3 The architecture of Masked Sentence Model based on BERT

Step1: Sentence Information Processing

In this step, for an abstract, each target sentence in the abstract of scientific papers is represented by the content of sentence.

For example, the second sentence (s_2) of the abstract shown in figure 1 is annotated with "Methods", the data format after this process step is shown in table 1. We use this representation to learn the internal features of the sentence.

Table 1 data format of sentence's content

Label	the content of the sentence
Methods	We selected the major journals (11 journals) collecting papers (more than 7,000) over the last five years from the top members of the research community, and read and analyzed the papers (more than 200) covering the topics.

Step2: Sentence Contextual Information Processing

In step2, we will get the contextual information of each sentence in the abstract. Here we adopt a new method that simply uses the whole abstract except the target sentence replacing by a [MASK] string to obtain the contextual information. And in this paper, we replace each word of the target sentence with 'aaa' characters to build the meaningless string of [MASK].

For the sentence(s_2), the length of this sentence is 37, so the data format after this step is shown in table 2 where the second sentence is replaced with 37 'aaa' characters. We input this contextual information into the BERT fine-tuning procedure and to learn the contextual features of the sentence.

Table 2 data format of sentence's context

Label	the context of the sentence

Methods	This survey aims at reviewing the literature related to Clinical Information Systems (CIS), Hospital Information Systems (HIS), Electronic Health Record (EHR) systems, and how collected data can be analyzed by Artificial Intelligence (AI) techniques. *aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa*. Then, we completed the analysis using search engines to also include papers from major conferences over the same five years. We defined a taxonomy of major features and research areas of CIS, HIS, EHR systems. We also defined a taxonomy for the use of Artificial Intelligence (AI) techniques on healthcare data. In the light of these taxonomies, we report on the most relevant papers from the literature. We highlighted some major research directions and issues which seem to be promising and to need further investigations over a medium- or long-term period.

Step3: MSM Integrate Processing

Thirdly, in this step, it integrates the content and contextual information of the sentence to construct the Masked Sentence Model. And it implements the integration by inputting the above two training samples together to the BERT fine-tuning procedure to train the Masked Sentence Model based on BERT-Base for move recognition task. For example, the final input representation for the second sentence (s_2) in this step is shown in table 3.

Table 3 data format of integrate sentence's content and context

Label	**the content & context of the sentence**
Methods	We selected the major journals (11 journals) collecting papers (more than 7,000) over the last five years from the top members of the research community, and read and analyzed the papers (more than 200) covering the topics.
Methods	This survey aims at reviewing the literature related to Clinical Information Systems (CIS), Hospital Information Systems (HIS), Electronic Health Record (EHR) systems, and how collected data can be analyzed by Artificial Intelligence (AI) techniques. *aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa aaa.* Then, we completed the analysis using search engines to also include papers from major conferences over the same five years. We defined a taxonomy of major features and research areas of CIS, HIS, EHR systems. We also defined a taxonomy for the use of Artificial Intelligence (AI) techniques on healthcare data. In the light of these taxonomies, we report on the most relevant papers from the literature. We highlighted some major research directions and issues which seem to be promising and to need further investigations over a medium- or long-term period.

3 Experiments & Results

3.1 Experiments design

In this paper, for different verification purposes, we design and conduct three different move recognition experiments based on BERT with the same neural network architecture and the same dataset. They only vary in the input data during the fine-tuning procedure.

Experiment 1 (or Exp1): An experiment based on the content of sentences. We fine-tune the BERT-Base model for downstream move recognition task by using the data format as shown in table 1, which only contains the content of sentences in fine-tuning input layer.

Experiment 2 (or Exp2): An experiment based on the context of sentences. The purpose of this experiment is exploring the rationality of our assumption based on distributional hypothesis and verifying the feasibility of the method of sentence context processing. This experiment was carried out based on the context of sentences by using the data format as shown in table 2 as BERT fine-tuning inputs.

Experiment 3(or Exp3): The most important experiment based on the MSM integrated information. In order to verify the effectiveness of the novel MSM model which is proposed in this paper. This experiment uses the data format as shown in table 3 which integrates the content and context of sentences as BERT fine-tuning inputs.

3.2 Data Sets

Our study evaluates the MSM model on the benchmark dataset PubMed 20K RCT (Dernoncourt and Lee, 2017) which contains approximately 20,000 medical scientific abstracts for sequential sentence classification. The dataset is based on the PubMed database of biomedical literature and each sentence of an abstract is labeled with its rhetorical role in the abstract using one of the following classes: Background, Objectives, Methods, Results, and Conclusions.

3.3 Hyper-parameters Setting

Our study uses the BERT-base model (Devlin et al., 2018) with a hidden size of 768, 12 Transformer blocks (Vaswani et al., 2017) and 12 self-attention heads. And fine-tunes with the following settings: a batch size of 5, a max sequence length of 512, the learning rate of 3e-5, the init_checkpoint of bert_base, the train steps of 100,000 and the warm-up steps of 10,000.

3.4 Evaluation Metrics

For each designed experiment, our study reports the performance with the evaluation metrics of precision (P), recall (R) and F1 score on the same test set provided by PubMed 20K RCT dataset (29,578 sentences from 2,500 abstract). The experimental results of each experiment are detailed in section 3.5.

3.5 Results

The Results of Exp1: based on the content of sentences

Table 4 shows the results of experiment based on the content of sentences. It can be found that just using the content of sentences can achieve effective results for move recognition at the average precision, recall, and F1 score of 86.75%, 86.61%, and 86.53%, respectively. In this experiment, it can get promising results on the categories of Methods and Results where both the precision and recall score are above 91%. On the category of Conclusions, particularly, the F1 score reaches to a relatively good score at 83.13%. However, for the categories of Background and Objectives, the F1 scores are 69.64% and 64.20% respectively, which can't satisfy our expectations very well and indicates a huge room for improvement.

Table 4 The results of Exp1: based on the content of sentences

Label	P	R	F1	Support
Background	64.37	75.85	69.64	3077
Objectives	73.55	56.97	64.20	2333
Methods	92.42	94.97	93.68	9884
Results	92.08	91.09	91.58	9713
Conclusions	84.95	81.38	83.13	4571
Avg / Total	86.75	86.61	86.53	29578

The Results of Exp2: based on the context of sentences

Table 5 presents the results of experiment based on the context of sentences, in which it only uses the context of sentences. The performance can reach the F1 score on average of 86.09%, and it indicates the validity of this method. These results prove that our basic assumption "sentences can be identified by their context in an abstract." is reasonable. And it shows that the method of using the context of sentences greatly improves the performance on the Background and Conclusions categories, whose average F1 performance improves 6.18% and 6.61% respectively, in comparison with the corresponding ones based on the content of sentences. However, the performance on the Methods and Results categories decreases slightly compared to the corresponding ones based on the content of sentences. These two experiments indicate that using the content of sentences performs better in Methods and Results categories and using the context of sentences performs better in Background, Conclusions and Objective categories.

Table 5 The results of Exp2: based on the context of sentences

Label	P	R	F1	Support
Backgroun	72.27	79.72	75.82	3077
Objectives	70.51	60.27	64.99	2333
Methods	90.70	89.80	90.25	9884
Results	87.71	89.20	88.45	9713
Conclusions	90.19	89.30	89.74	4571
Avg / Total	86.13	86.15	86.09	29578

The Results of Exp3: based on MSM integrated information

Table 6 shows the BERT's fine-tuning experimental results based on MSM model with integrated information of content and context of sentences. It shows that the novel model achieves the best results in the overall experiments, and the average precision, recall and F1 score can reach 91.22%, 91.30%, and 91.15% respectively. Compared to the first two experiments, the results of this experiment increase by 4.62% and 5.06% respectively in terms of F1 score. The results show that the novel Masked Sentence Model performs better performance on the task of move recognition. Furthermore, the study verifies that the model can effectively improve the performance by incorporating the content and context information of sentences.

Table 6 The results of Exp3: based on MSM integrated information

Label	P	R	F1	Support
Background	75.26	81.18	78.11	3077
Objectives	78.08	61.98	69.10	2333
Methods	92.98	97.48	95.17	9884
Results	96.02	93.74	94.87	9713
Conclusions	94.70	94.51	94.60	4571
Avg / Total	91.22	91.30	91.15	29578

3.6 Result Analysis

Table7 shows the comparing results of the above-mentioned three experiments, which better explain the advantage of our integrated MSM model.

Table 7 Compared results of the experiments

Label	Exp1	Exp2	Exp3	Exp3-Exp1	Exp3-Exp2
	F1	F1	F1	+F1	+F1
Background	69.64	75.82	78.11	**8.47**	2.29
Objectives	64.20	64.99	69.10	**4.9**	4.11
Methods	93.68	90.25	95.17	1.49	**4.92**
Results	91.58	88.45	94.87	3.29	**6.42**
Conclusions	83.13	89.74	94.60	**11.47**	**4.86**
Avg / Total	86.53	86.09	91.15	4.62	5.06

From the “Exp3-Exp1”, it can be found that the MSM model greatly improves the performance on the ‘Conclusions’ and ‘Background’ categories, whose average F1 performance improve 11.47% and 8.47% respectively, in comparison with the corresponding ones based only on the content of the sentence. Then followed by the 4.9% improvement on ‘Objectives’ category. The impact on the ‘Methods’ and ‘Results’ categories is relatively small. Based on these comparisons, it indicates that the ‘Conclusions’ and ‘Background’ categories are more context-sensitive and can achieve huge improvements by incorporating contextual information. This makes sense because the ‘Background’ move always appears at the beginning of the abstracts and the ‘Conclusions’ move appears at the end of the abstract. So by adding contextual information to the input, the model will learn the positional information to a certain extent.

Correspondingly, from the “Exp3-Exp2”, in comparison with the corresponding ones based only on the context of the sentence, the MSM model got the highest F1 value growth on the ‘Results’ category(6.42%), the lowest F1 value growth on the Background category (2.29%) and a relative balanced growth in the other three categories. This indicates that the content of the sentence also plays an important role to identify the rhetorical role of the sentence, which used to express the author's writing intention.-

4. Comparisons & Discussion

In this section, we compare the model proposed in this paper with other models, and discuss the compared results. Table 8 lists our model and the other experimental models evaluated on the PubMed 20k corpus. MaskedSentenceModel based on BERT presented in this paper is denoted by "Our Model". "Others" contains HSLN model, BERT-Base model, and two models of SciBERT.

It shows that HSLN's average F1 score still ranks the first based on Bi-LSTM+CRF methods. Our model achieves better performance than the current models based on BERT, which outperforms 4.34 points than SciBERT (BaseVocab), and 4.96 points than BERT-Base. However, our current model is still 1.45 points lower than HSLN. This makes sense because the sequence tag information considered in the HSLN model is not considered in our model, so there is room for improvement based on BERT.

Table 8 PubMed 20k RCT results

	Models	F1(PubMed 20k RCT)
Our Model	**MaskedSentenceModel_BERT**	**91.15**
Others	**HSLN-RNN(Jin and Szolovits, 2018) (SOTA)**	**92.6**
	BERT-Base (Beltagy et al.,2018)	86.19
	Sci BERT (SciVocab) (Beltagy et al.,2018)	86.80
	Sci BERT (BaseVocab) (Beltagy et al.,2018)	86.81

5. Conclusions & Future Work

This paper presents a novel approach to recognizing moves in scientific abstracts using Masked Sentence Model based on BERT. It demonstrates that integrating content and context information of sentences by using MSM model to learn the internal and contextual features of sentences can improve overall performance of recognitions. The proposed method achieves more successful results than other previous BERT based methods. It outperforms the BERT-Base and SciBERT results by 4.96% and 4.34% respectively on the public dataset PubMed 20k RCT. Because the model doesn't take sequential features of move labels into account, HSLN-RNN still has better performance than it.

Our MSM approach is general and easy to apply. It can achieve a better performance improvement on the move recognition task just by making some optimizations in the input layer without any change on the BERT's internal structure of neural networks. And we believe that our model is effective for many other context-sensitive NLP tasks including text classification, sentiment analysis.

Although our model are proved to have great performance, there is also some limitations. Our current method is still relatively simple because we simply replace each word of the sentences with a meaningless string 'aaa' to generate the contextual representation of the sentences. In the future, we will improve the performance of the MSM method from the following aspects:

1. We will try to modify the method of generating the contextual information of the sentences such as using a fixed length string of 30 'aaa' to mask the target sentences and analyze its effect.

2. We plan to extend our MSM to cover many other important features for move recognition such as sequential features which are not incorporated in our MSM approach.

3. Also we would like to try modifying the structure of neural networks to fit the special input layer proposed in this study. In that way, this context-sensitive approach could be more efficient.

Acknowledgments

This work is supported by the project "The demonstration system of rich semantic search application in scientific literature" (Grant No. 1734) from the Chinese Academy of Sciences.

Author Contributions

Zhixiong Zhang (zhangzhx@mail.las.ac.cn) and Gaihong Yu(yugh@mail.las.ac.cn) designed and produced the research. Huan Liu(liuhuan@mail.las.ac.cn) conducted the experiments. Gaihong Yu wrote the main body of the paper. Zhixiong Zhang done a lot of modifications and improvements and finally completed the paper. Huan Liu and Liangping Ding(dingliangping@mail.las.a.cn) done a lot of paper revision and improvement work especially in the Introduction and Methodology Section.

References

Amini I, Martinez D, Molla D. Overview of the ALTA 2012 shared task[J]. 2012.

Badie K, Asadi N, Tayefeh Mahmoudi M. Zone identification based on features with high semantic richness and combining results of separate classifiers[J]. Journal of Information and Telecommunication, 2018: 1-17.

Basili R, Pennacchiotti M. Distributional lexical semantics: Toward uniform representation paradigms for advanced acquisition and processing tasks[J]. Natural Language Engineering, 2010, 16(4): 347-358.

Beltagy I , Cohan A , Lo K . SciBERT: Pretrained Contextualized Embeddings for Scientific Text[J]. 2019.

Dasigi P, Burns G A P C, Hovy E, et al. Experiment segmentation in scientific discourse as clause-level structured prediction using recurrent neural networks[J]. arXiv preprint arXiv:1702.05398, 2017.

Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding [J]. arXiv preprint arXiv:1810.04805, 2018.

Ding Liangping, Zhang Zhixiong,Liu Huan.Research on Factors Affecting the SVM Model. Performance on Move Recognition[J].Data Analysis and Konwledge Discovery,2019

Firth J R. A synopsis of linguistic theory, 1930-1955[J]. Studies in linguistic analysis, 1957.

Fisas B, Ronzano F, Saggion H. A multi-layered annotated corpus of scientific papers[C]//Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016). 2016: 3081-3088.

Franck Dernoncourt and Ji Young Lee. 2017. Pubmed 200k rct: a dataset for sequential sentence classification in medical abstracts. In IJCNLP.

Gerlach M, Peixoto T P, Altmann E G. A network approach to topic models[J]. Science advances, 2018, 4(7): eaaq1360.

Hirohata K, Okazaki N, Ananiadou S, et al. Identifying sections in scientific abstracts using conditional random fields[C]//Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-I. 2008.

Mingbo Ma, Liang Huang, Bing Xiang, and Bowen Zhou. 2015. Dependency-based convolutional neu- ral networks for sentence embedding. arXiv preprint arXiv:1507.01839.

Peters M E, Neumann M, Iyyer M, et al. Deep contextualized word representations[J]. arXiv preprint arXiv:1802.05365, 2018.

Radford A, Narasimhan K, Salimans T, et al. Improving language understanding by generative pre-training[J].URL:https://s3-us-west-2.amazonaws.com/openai-assets/researchcovers/languageunsupervised/language understanding paper. pdf, 2018.

Siwei Lai, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Recurrent convolutional neural networks for. text classification. In AAAI, volume 333, pages 2267– 2273.

Swales,J.M. Research Genres: explorations and applications [M].Cambridge:Cambridge University. Press,2004:228-229.

Taylor W L. “Cloze procedure”: A new tool for measuring readability[J]. Journalism Bulletin, 1953, 30(4): 415-433.

TEUFEL S. Argumentative zoning:information extraction from scientific text[D]. Edinburgh:University of Edinburgh,1999.

Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Advances in neural information. processing systems. 2017: 5998-6008.

Yamamoto Y, Takagi T. A sentence classification system for multi-document summarization in the. biomedical domain [C]//Proceedings of International Workshop on Biomedical Data Engineering. 2005: 90-95.

Yoon Kim. 2014. Convolutional neural net- works for sentence classification. arXiv preprint. arXiv:1408.5882.

Zhang Z, Liu H, Ding L, et al. Moves Recognition in Abstract of Research Paper Based on Deep Learning[C]//2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL). IEEE, 2019: 390-391.